

# The Forward Problem of Spectral Reflection Prediction: Mutual Match between Framework Selection and the Training Set Volume

Dmitry A. Tarasov<sup>1,a)</sup>, Oleg B. Milder<sup>1,b)</sup>

<sup>1</sup>*Ural Federal University, Mira str., 19, Ekaterinburg, RUSSIA 620002*

<sup>a)</sup>Corresponding author: [datarasov@yandex.ru](mailto:datarasov@yandex.ru)

<sup>b)</sup>[milder@mail.ru](mailto:milder@mail.ru)

**Abstract.** Novel technology of the color reproduction is closely related to a variety of color prediction techniques. Along with the deterministic models, those based on the use of artificial neural networks have recently begun to appear. An important problem in the application of a neural network approach is the choice of the training set and network training algorithms. In our work, we describe the results of a computational experiment where various configurations of the artificial neural networks along the different volumes of training subsamples were simultaneously determined providing a satisfactory accuracy of the spectral reflection prediction. The results show that each subsample might be mapped to a particular network configuration.

**Keywords:** Color reproduction, Spectral reflection prediction, Artificial neural networks, Framework, Training set.

## INTRODUCTION

The purpose of color management systems is to calibrate the color response of the reproduction system by combining a set of the initial dyes. In today's printing, the traditional subtractive model using 4 dyes (CMYK) is common. The color prediction models can help the color management system software determine which ink set and how to select and mix to create a specific color on a specific substrate using a specific ink. The color modeling is inextricably linked with the prediction of the reflection spectrum of a colored sample. Often, the color prediction and spectrum prediction are used interchangeably. In a generalized form, the forward problem of spectral reflection prediction is formulated as a forecast with a given accuracy of the spectral response of a colored sample in accordance with a specific formulation of the initial dyes.

To date, a large number of color prediction models have been developed. Empirical surface models take into account the imposition of the ink halftones and do not take into account the propagation of the light and its attenuation in the print. Physically inspired models include a more detailed analysis of the light-in-finger interaction based on a prediction of how light paths go and what the resultant attenuation is. The ink distribution models estimate the effective surface of the ink dot after it is printed at a given nominal surface coverage compared to the given characteristics of this dot. Models for predicting the color of spectral reflection deal with probabilistic models of light propagation in a sample and study the influence of various factors that effect on the range of printed colors [1]. In these models, the tone reproduction curves (TRC) of several initial colorants are utilized to characterize the physical enhancement of the dotted halftones of ink on the substrate and under all conditions of solid ink overlap [2–4].

The color prediction models are successfully applied to color reproduction management [5–9]. However, growth of the forecasts accuracy comes at a price. Thus, the requirement for computational power in the numerical solution of direct forecast problems grows exponentially with increasing the model complexity, since  $n$  dyes require solving a system of  $n^2$  equations. Consequently, these approaches are hardly applicable to the real-time control systems.

Moreover, since most of the calculations are based on empirical relationships, the accuracy of the prediction for estimating the reflection spectrum of certain shades remains relatively low.

One of the promising ways to build a high-speed color prediction model is to replace the deterministic approach with a statistical one, in particular, the use of artificial neural networks. This approach is suitable for solving complex nonlinear problems and has repeatedly proved its efficiency, speed and undemanding of computing resources. Preliminary results [10] showed a wide prospect of using the ANN in color prediction, in particular, the applicability of the Bayesian regulated ANN was proved [11].

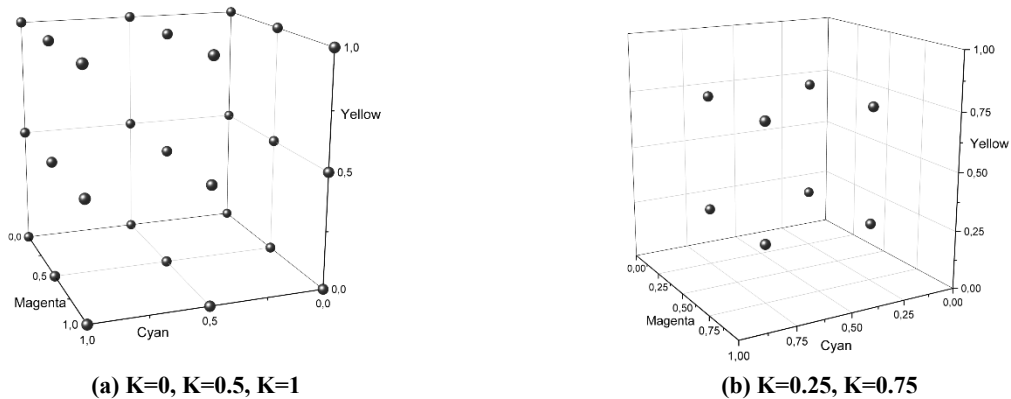
Preliminary studies [12] showed that the main difficulty in using the neural network approach is the selection of the network configuration and determining the size of the training set. This work is devoted to development of the ANN technique for direct spectral reflection prediction. We establish a relationship between the framework selection and the training set volume.

## APPROACH AND EXPERIMENTAL

For the experiment, we use the electrophotographic (4-colors, CMYK) printer Konica-Minolta C6000L. Print mode was 1200×1200 dpi. Substrate was the coated paper Murim NeoStar 120 g/m<sup>2</sup>. The measurement tools: spectrophotometer X-Rite iOne iSis + X-Rite ProfileMaker package. Charts generation is made in the ArgyllCMS package. For the ANN development, we have applied Matlab 16. Quantitative assessment of the model prediction accuracy was carried out by the common  $dE_{2000}$  color difference formula.

Based on the preliminary studies, we have chosen an ANN type. It is a fully connected multilayer perceptron with one hidden layer. The training techniques are the Levenberg–Marquardt method with the Bayesian regularization [11].

One of the key issue in the neural networks modeling is a training set selection. We have chosen a sample set of colors that are evenly distributed over the entire printing system color body. Training sample sets form regular grids in the colorants recipe space (Fig. 1a). The selection points are also located in the centers of the main cells (Fig. 1b). The number of grid nodes per each formal CMYK direction is determined by the dimension of the abstract cube in the color space that is called the “bodycube”. Fig. 1 shows the complete organization of a training sample of the bodycube of dimension 3 (bc3). In the experiment, four bodycubes (bc3–bc6) have been utilized (see Table 1).



**FIGURE 1.** Color bodycube of dimension 3 (bc3) in the CMY color space; the black channel (K) fractions are indicated below.

**TABLE 1.** ANN training sets configurations.

Training set	bc6	bc5	bc4	bc3
Number of patches	1921	881	337	97
Dimension, R	6	5	4	3
Step of the main grid, tone fraction	0.20	0.25	0.33	0.50

The whole experiment approach resembles that from [12]. In order to avoid the appearance of the negative

values of reflection coefficients in the predictions, we carry out all the calculations regarding the spectral density  $D_i(\lambda) = -\log_{10}\rho_i(\lambda)$  with further recalculation back into the spectral reflectance  $\rho_i(\lambda) = 10^{-D_i(\lambda)}$  (where  $\lambda$  is the wavelength in the range (380–730) nm,  $i$  is the counter of patches to print).

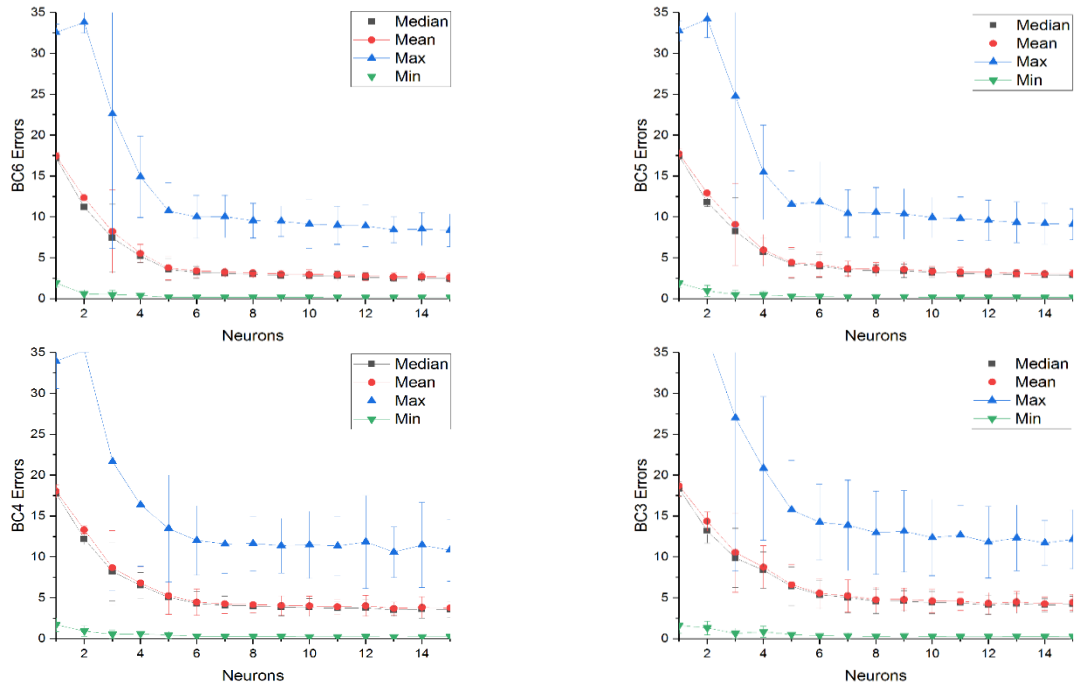
Since the main task of this work is to establish the relationship between the framework and training set volume, we have varied the number of the hidden neurons as well as the training sets. The number of hidden neurons vary from 1 to 15. The ANN output is the predicted spectrum. For each spectra, the CIE Lab coordinates are calculated. The test chart is then measured and the actual values of CIE Lab coordinates are established. The color difference between predicted and measured values for each patch of the test chart is calculated by  $dE_{2000}$  equation.

Such, in the experiment, we have predicted the spectral reflectance (36 spectral zones from 380 to 730 nm, with the step 10 nm) by test chart patches recipes (4 values corresponding to CMYK percentage in the layout). Thus, the configuration of the network is  $4 \times N \times 36$ , where  $N=1, 2, \dots, 15$  is the number of neurons in the hidden layer. We have combined all frameworks with all training sets and applied multiple (51 times) computations to each pair in order to assess the  $dE_{2000}$  distributions.

The final stage of the experiment is application of the trained networks to the spectral predictions. We feed each of the trained networks with data of the independent validation recipes set that contains 2790 patches randomly distributed over the printing device color space. This scale is built with a particular limit: no less than a half of the patches must be located near the achromatic axis. Errors between the factual and predicted spectra are evaluated by the  $dE_{2000}$  equation. We have also assessed mean, median, min, and max values of the errors.

## RESULTS AND DISCUSSION

In the experiment, we have created and trained 3060 different artificial neural networks. This value is formed by 4 training sets multiplied by 15 frameworks and 51 repetitions. Mean, median, min, and max values of the prediction errors are shown in Fig. 2.



**FIGURE 2.** Errors ( $dE_{2000}$ ) with the SD intervals for different training sets (bcX).

As it is seen from the figures, both growth in the number of the hidden neurons and increase in the training set volume lead to decrease in the prediction error ( $dE_{2000}$ ). If we take the value of  $dE_{2000}$  equal to 5 for the level of indistinguishability of color differences, then the minimal sample bc3 provides an acceptable accuracy with 8 neurons in the hidden layer. Further, as the sample increases, the required number of neurons decreases: 6 neurons

for bc4, 5 neurons for bc4 and 4 neurons for bc3. Thus, it can be argued that for an arbitrary sample size, the network configuration when solving the direct spectral prediction problem is of 6–8 hidden neurons.

Further, we described the dependence of the median error ( $dE_{2000}$ ) on the number of hidden neurons using the decreasing exponent of the form  $dE_m(n) = dE_\infty + A * \exp(-n/n_0)$ , where  $n$  is the number of neurons in the hidden layer;  $dE_m$  is the median value of the forecast error expressed as the  $dE_{2000}$ ;  $A$  is an abstract amplitude value;  $n_0$  is the exponent coefficient equal to the number of the hidden neurons, the addition of which reduces the median error by a factor of  $\exp$ ;  $dE_\infty$  is the hypothetical value of the prediction error median in the case of an infinite number of hidden neurons. The simulation results are shown in Table 2. Exponential fitting resulted in the  $dE_\infty$  approx. 2.5. Thus, within the proposed model, it is impossible to make a forecast with an error less than 2.5 units of  $dE_{2000}$ .

**TABLE 2.** Exponential fitting features.

Training set	bc6	bc5	bc4	bc3
Number of patches	1921	881	337	97
$n_0$	1.85±0.04	1.99±0.04	1.95±0.04	2.23±0.06
$dE_\infty$	2.58±0.05	2.95±0.03	3.59±0.05	4.15±0.05

## CONCLUSION

The complicated nonlinear direct problem of spectral reflection prediction might be successfully solved by mean the artificial neural networks. They proved to be a convenient modeling tool, undemanding to computational resources and giving high accuracy of prediction. One of the key task in the approach is to define the framework structure and training set volume.

In our experiment, we utilized the feed-forward network of the multilayer perceptron with one hidden layer and the Bayesian regularization. The study aimed to establish a relationship between the number of the hidden neurons and the training set volume. Four training sets were formed from the constructed bodycubes from the color space of the printing system. The number of neurons also varied from 1 to 15.

We showed that such a framework would not exceed the precision of prediction better than  $dE_{2000}$  2.5 units. Anyway, this result might be considered satisfactory. In order to provide the accuracy of 5  $dE_{2000}$  units, one should apply the model with 6–8 hidden neurons for any training set volume.

## REFERENCES

1. R. Bala, “Device characterization” in *Digital Color Imaging Handbook*, ed. G. Sharma (CRC Press, Boca Raton, FL, 2003), pp. 269–379.
2. R. Balasubramanian, *J. Electronic Imaging*, **8**(2), 156–166 (1999).
3. R.D. Hersch and F. Cr   , “Improving the Yule–Nielsen modified spectral Neugebauer model by dot surface coverages depending on the ink superposition conditions” in *Proc. SPIE 5667* (2005), pp. 434–445.
4. D.R. Wyble and R.S. Berns, *Col.Res.&App.*, **25**, 4–19 (2000).
5. N.P. Garg, A.K. Singla, and R.D. Hersch, *J.Imag.Sci.Tech* **52**(4), 040908–040908-5 (2008).
6. J.S. Arney, P.G. Engeldrum, and H. Zeng, *J.Imag.Sci.Tech* **39**, 502–508 (1995).
7. S. Livens, “Optimisation of Printer Calibration in the Case of Multi Density Inks” in *Conference on Color in Graphics, Imaging, and Vision*, CGIV 2002. Final Program and Proceedings (2002), pp. 633–638.
8. L. Chagas, A. Blayo, and P. Giraud “Color Profile: methodology and influence on the performance of ink-jet color reproduction” in *IS&T's NIP20. International Conference on Digital Printing Technologies* (2004), pp. 655–659.
9. Y-J. Wu Reducing “Ink-jet Ink Consumption with RIP software for POP Display Media. Digital Fabrication and Digital Printing” in *NIP30 Technical Program and Proceedings* (2014), pp. 108–111.
10. O.B. Milder and D.A. Tarasov “Spectral reflection prediction by artificial neural network” in *CEUR Workshop Proceedings 2076* (2018) pp. 86–95.
11. M. Kayri *Math.Comp.App.* **21**(2), 21020020 (2016)
12. O. Milder, D. Tarasov, and A. Tyagunov “The Artificial Neural Network Structure Selection Algorithm in the Direct Task of Spectral Reflection Prediction” in *WSEAS Trans. on Systems and Control* **14** (2019) pp. 65–70.